I. PURCARU
Academy of Economic Studies

On Some Measures of Dependence between Random Variables

This paper presents a connection between two measures of dependence between random variables followed by two generalizations of the dependence, and together with them, by some extensions of Rényi's postulates ([7] [8]).

1. INTRODUCTION

Let us consider a certain phenomenon or socio-economic process, whose evolution in time, is determined by the influence of a number of factors. We suppose the degree of influence of each factor on the studied process to be known. It seems interesting to know the manner how the considered factors act upon the studied phenomenon, namely, whether they act dependently or independently, between them.

There are sufficient indicators able to answer to this problem — more or less satisfactorily. One of these indicators is the classical correlation coefficient, which — notwithstanding its disadvantages (see [4], [6], [7], [8]) is the most frequently utilized. We shall present further down also other indicators of the interdependence between these and the considered factors, and afterwards we shall show also which relation is there between them and the classical correlation coefficient. We shall present finally also a generalization of those indicators and, consequently, of the connection between them. We shall admit that, at a given moment, each factor is described by a random variable and, consequently, the considered factors system will be described by a random vector.

2. INFORMATIONAL INDICATORS OF INTERDEPENDENCE

Let us consider at beginning a bidimensional random vector $X = (X_1, X_2)$ whose components are continuous or discrete random variables, with finite entropy. We shall consider for the moment that the random variables are discrete, so as to better see the expressions and the practical mode of utilization of the presented indicators. Hence

$$X_k : \begin{pmatrix} x_k^{i_k} \\ p(x_k^{i_k}) \end{pmatrix} i_k \in I_k, \; k = 1, 2 \text{ and } X : \begin{pmatrix} x_1^{i_1} x_2^{i_2} \\ p(x_1^{i_1} x_2^{i_2}) \end{pmatrix} \tag{1}$$

and consequently, the attached entropies will be:

$$H(X_k) = -\sum_{i_k \in I_k} p(x_k^{i_k}) \log p(x_k^{i_k}) \text{ and } H(X) =$$

$$= -\sum_{i_1 \in I_1} \sum_{i_2 \in I_2} p(x_1^{i_1}, x_2^{i_2}) \log p(x_1^{i_1}, x_2^{i_2}). \tag{2}$$

Let us consider now magnitude $W(X; X_1, X_2)$ defined by relation

$$W(X; X_1, X_2) = H(X_1) + H(X_2) - H(X). \tag{3}$$

Theorem 1. Magnitude $W(X; X_1, X_2)$ defined by relation (3), satisfies the following properties

1) $0 \leqslant W(X; X_1, X_2) < \infty$

2) $W(X; X_1, X_2) = W(X; X_1, X_2)$

3) $W(X; X_1, X_2) = 0$ if and only if X_1 and X_2 are probabilistically independent;

4) if $\varphi_1(x)$ and $\varphi_2(x)$ are two real functions, defined biunivocally on the real axis, then

$$W(X; \varphi_1(X_1), \varphi(X_2)) = W(X; X_1, X_2)$$

5) $W(X; X_1, X_2) \leqslant min\,[H(X_1), H(X_2)] \leqslant max\,[H(X_1), H(X_2)]$

6) if X_1 is completely dependent on X_2, then

$$W(X; X_1, X_2) = H(X_1)$$

7) if X_1 and X_2 are completely dependent from one another

$$W(X; X_1, X_2) = H(X_1) = H(X_2) = H(X_1, X_2)$$

and reciprocally

8) $W(X; X_1, X_2) \leqslant \frac{1}{2}[H(X_1) + H(X_2)] \leqslant H(X_1, X_2)$

with equality if and only if X_1 and X_2 are completely dependent from one another.

The proof of the theorem is immediate. Variable X_1 is completely dependent on variable X_2 if conditioned entropy $H(X_1/X_2) = 0$, and random variables X_1 and X_2 are completely dependent from one another, if $H(X_1/X_2) = H(X_2/X_1) = 0$.

Magnitude $W(X; X_1, X_2)$ defined by relation (3) is called measure of interdependence or cohesion existing between random variables X_1 and X_2. We can do various normings of measure $W(X; X_1, X_2)$,so as to be able to compare more easily the various values the measure could take [4]. One of the se normings, denoted by $r(X_1, X_2)$, is given by relation

$$r(X_1, X_2) = \frac{W(X; X_1, X_2)}{H(X_1, X_2)} \tag{4}$$

and it is called informational correlation coefficient or interdependence coefficient (see [4], [5], [6]).

We have, as a consequence of theorem 1:

Theorem 2. Interdependence coefficient $r(X_1, X_2)$, defined by relation (4), satisfies the properties:

1) it is defined for any random variable pair X_1 and X_2, none of them being constant with probability 1

2) $0 \leqslant r(X_1, X_2) \leqslant 1$

3) $r(X_1, X_2) = 0$ if and only if X_1 and X_2 are probabilistically independent.

4) if $\varphi_1(x)$ and $\varphi_2(x)$ are two real functions biunivocally defined on the real axis, then

$$r(\varphi_1(X_1), \varphi_2(X_2)) = r(X_1, X_2)$$

5) $r(X_1, X_2) = 1$ if and only if X_1 and X_2 are completely dependent from one another.

In order to be aware of the practical expression of this coefficient and especially of its mode of utilization, let us suppose that random variables X_1 and X_2 take n_1 and respectively n_2 distinct values, and let us denote by

$$\begin{aligned} p_{ij} &= P(X_1 = x_1^i,\ X_2 = x_2^j),\ 1 \leqslant i \leqslant n_1,\ 1 \leqslant j \leqslant n_2 \\ p_{i\cdot} &= P(X_1 = x_1^i) = \sum_{j=1}^{n_2} p_{ij},\ 1 \leqslant i \leqslant n_1 \\ p_{\cdot j} &= P(X_2 = x_2^j) = \sum_{i=1}^{n_1} p_{ij},\ 1 \leqslant j \leqslant n_2 \end{aligned} \tag{5}$$

the corresponding probabilities which would appear in formula (4). becomming

$$r(X_1, X_2) = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} p_{ij} \log \frac{p_{i\cdot}\ p_{\cdot j}}{p_{ij}}}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} p_{ij} \log p_{ij}}. \tag{6}$$

If we replace in expression (6) the probabilities with the relative frequencies appearing, following n observations, then the working shape of coefficient $r(X_1, X_2)$ becomes:

$$r(X_1, X_2) = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} f_{ij} \log \frac{f_i \cdot f \cdot j}{f_{ij}} - n \log n}{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} f_{ij} \log f_{ij} - n \log n} \tag{7}$$

where f_{ij} represents the absolute frequency of pair (x_1^i, x_2^j) and f_i. and $f \cdot j$, the absolute frequences of values x_1^i and x_2^j, respectively.

It is obvious that, we can easily construct a program for computer for the practical computations of $r(X_1, X_2)$, according to relation (7), if there are, of course, many data.

We can make immediately generalizations for several random variables or random vectors starting from measure $W(X; X_1, X_2)$ and $r(X_1, X_2)$, respectively (see [3], [4], [5], [6]). Let consequently be random vector $X = (X_1, X_2, ..., X_N)$. A first generalization (see [3], [4]) was done by considering expression

$$W(X_i, X_1, X_2, ... X_N) = \sum_{i=1}^{N} H(X_i) - H(X) \tag{8}$$

called measure of interdependence between the N random variables, and which satisfies properties 1), 2), 3), 4) and 7) from theorem 1, adapted to generalization .As a consequence of the last property we introduced also the coefficient of total interdependence (see [4]).

$$R(X_1, X_2, ..., X_N) = \frac{W(X_i\ X_1, X_2, ..., X_N)}{(N-1)\ H(X)} \tag{9}$$

A second generalization (see [4]) was done by considering M random vectors, fromed with the N components of X, denoted by $Y_1, Y_2, ..., Y_M$, each vector Y_j, $1 \leqslant j \leqslant M$, having at least one component. In this case, the interdependence has the same form as (8), namely:

$$\overline{W}(X; Y_1, Y_2, ..., Y_M) = \sum_{j=1}^{M} H(Y_j) - H(X) \tag{10}$$

and by analogy with (9), the interdependence coefficient will be:

$$\overline{R}(Y_1, Y_2, ..., Y_M) = \frac{\overline{W}(X; Y_1, Y_2, ..., Y_M}{(M-1)\ H(X)}. \tag{11}$$

A third generalization of magnitudes $W(X; X_1, X_2)$ and $r(X_1, X_2)$ (see [4]) is represented by the particularization of relation (10) and (11), taking $M = 2$, which leads to the measure of interdependence between two random vectors. A dummy-table of gathering observations [6], which generalizes the classical bidimensional contingency, table was given, in order to use easily these indicators, taking into account the explanations connected to relation (7).

3. CONNECTION BETWEEN THE MEASURE OF INTERDEPENDENCE AND THE CLASSICAL CORRELATION COEFFICIENT

Though both magnitudes are very different from one another, both as form and especially from the viewpoint of their content (see [4]), there are certain cases when there are also closenesses between them. We shall suppose, further down that random vector $X = (X_1, X_2)$ has a continuous distribution and we shall denote by $p(x_1, x_2)$ the corresponding distribution density and by $p(x_1)$ and $p(x_2)$ the marginal distribution densities, and no danger of confession is possible.

Let us assume that

$$p(x_1, x_2) = \frac{1}{2\pi}\sqrt{ab - h^2}\,\exp\left\{-\frac{1}{2}[a(x_1 - \lambda_1)^2 + \right.$$

$$\left. + 2h(x_1 - \lambda_1)(x_2 - \lambda_2) + b(x_2 - \lambda_2)^2]\right\} \tag{12}$$

where $a > 0$, $b > 0$ an $ab - h^2 > 0$, and λ_1 and λ_2 are two real constants. We shall consider further down, without diminishing in no way the generality of the problem, that $\lambda_1 = \lambda_2 = 0$ which implies

$$p(x_1, x_2) = \frac{1}{2\pi}\sqrt{ab - h^2}\,\exp\left\{-\frac{1}{2}(ax_1^2 + 2hx_1x_2 + bx_2^2)\right\}. \tag{13}$$

Theorem 3. The marginal distributions of random vector $X = (X_1, X_2)$ with common distribution (13) are given by

$$p(x_1) = \sqrt{\frac{\alpha}{\pi}}\,e^{-\alpha x_1^2}, \quad \alpha = \frac{ab - h^2}{2b} \tag{14}$$

$$p(x_2) = \sqrt{\frac{\beta}{\pi}}\,e^{-\beta x_2^2}, \quad \beta = \frac{ab - h^2}{2a}.$$

The proof is immediate. We have also:

Theorem 4. Correlation coefficient $\rho(X_1, X_2)$ and interdependence measure $W(X; X_1, X_2)$ are given by relations:

$$\rho(X_1, X_2) = -\frac{h}{\sqrt{ab}}$$

$$W(X_i; X_1, X_2) = \frac{1}{2}\log\frac{ab}{ab - h^2}. \tag{15}$$

The following theorem ensures the existence of a connection between both measures of dependence which, in case of distribution (13) have the expressions given by (15).

Theorem 5. Let $X = (X_1, X_2)$ be a random vector, $\rho(X_1, X_2)$ the classical correlation coefficient, and $W(X; X_1, X_2)$ the measure of interdependence between X_1 and X_2 defined by relation (3). There is always a function $\varphi: R_+ \to [0, 1]$ such that, denoting by

$$\delta(X_1, X_2) = \varphi(W(X; X_1, X_2)) \tag{16}$$

another measure of the dependence between random variables X_1 and X_2, the following possibilities be satisfied:

1) $\delta(X_1, X_2)$ is defined by any pair (X_1, X_2) none of the variables being constant with probability 1.

2) $\delta(X_1, X_2) = (X_2, X_1)$ i.e. it is a function symmetrical in both variables.

3) $0 \leqslant \delta(X_1, X_2) \leqslant 1$ for any $X = (X_1, X_2)$.

4) $\delta(X_1, X_2) = 0$ if and only if X_1 and X_2 are probabilistically independent.

5) $\delta(X_1, X_2) = 1$ if $X_1 = \eta_1(X_2)$ or $X_2 = \eta_2(X_1)$.

6) if ξ_1 and ξ_2 are two real functions biunivocally defined on the real axis, then

$$\delta(\xi_1(X_1), \xi_2(X_2)) = \delta(X_1, X_2).$$

7) if the random vector $X = (X_1, X_2)$ has the distribution (13), then

$$\delta(X_1, X_2) = |\rho(X_1, X_2)|.$$

Proof. Let us consider function

$$\varphi^2(x) = 1 - e^{\mu x + \nu}, \quad \mu \text{ and } \upsilon \text{ real constant.} \tag{17}$$

We notice at once that, by putting $x = W(X; X_1, X_2)$ and prescribing to function $\varphi(x)$ to verify the above properties, we find $\mu = -2$ and $\upsilon = 0$, which leads to function

$$\delta(X_1, X_2) = \varphi(W(X; X_1, X_2)) = \sqrt{1 - e^{-2W(X; X_1, X_2)}}. \tag{18}$$

By analysing the properties possed by measure of dependence $\delta(X_1, X_2)$ and by taking into account Rényi's postulates concerning the definition of a better measure of dependence between the random variables (see [2], [7], [8]), we may consider $\delta(X_1, X_2)$ the informational correlation coefficient as a good indicator of the dependence between X_1 and X_2.

This informational correlation coefficient leads us to the introduction of a measure of interdependence between random variables more general and more powerful than the classical correlation coefficient, and which, for certain distributions coincides with the latter. The distribution family,for which we have identity between both measures, is represented by relations (12) or (13), where a, b and h being real parameters, the family includes a triple infinity of such distributions.

The normal bidimensional distribution represents a distribution subfamily of family $\{p(x_1, x_2; a, b, h)\}$ given by (13).

As a matter of fact, putting in relation (13)

$$a = \frac{1}{\sigma_1^2(1-\rho^2)}, \quad b = \frac{1}{\sigma_2^2(1-\rho^2)} \text{ and } h = -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \tag{19}$$

where σ_1^2 and σ_2^2 are the dispersions of variables X_1 and X_2, and ρ the correlation coefficient, we shall become just the density of normal bidimensional distribution.

Thus, we have a rich distribution family, for which the introduced informational measure coincides with the classical correlation coefficient.

4. TWO GENERALIZATIONS OF THE CONNECTION BETWEEN THE MEASURE OF INTERDEPENDENCE AND THE CLASSICAL CORRELATION

Let us suppose that the random N-dimensional vector $X = (X_1, X_2, \ldots, X_N)$ has the distribution density

$$p(x_1, x_2, \ldots, x_N) = \frac{1}{(2\pi)^{N/2}} \sqrt{|A|}\, e^{-\frac{1}{2}(x-\lambda)A(x-\lambda)} \tag{20}$$

where $A = (aij)$ with $aii > 0$ for all $1 \leqslant i, j \leqslant N$ and $|A| > 0$ and λ is a real constant vector. In the following formulas, for ease in writing and calculation, we shall take $\lambda = 0$, and consequently we shall work with distribution

$$p(x_1, x_2 \ldots x_N) = \frac{1}{(2\pi)^{N/2}} \sqrt{|A|}\, e^{-\frac{1}{2}xAx'}. \tag{21}$$

Theorem 6. The marginal distribution densities corresponding to distribution (21) of vector X are given by

$$p(x_i) = \sqrt{\frac{|A|}{2\pi M_{ii}}}\, e^{-\frac{1}{2}\frac{|A|}{M_{ii}}x_i^2}, \quad 1 \leqslant i \leqslant N \tag{22}$$

where M_{ii} is the minor of element a_{ii} from matrix A, $M_{ii} > 0$. Relation (22) is obtained directly, through calculation.

Theorem 7. If random vector X has distribution (21), then

$$H(X_i) = \frac{1}{2}\log\frac{2\pi M_{ii}}{|A|}, \quad 1 \leqslant i \leqslant N$$

$$H(X) = \frac{1}{2}\log\frac{(2\pi)^N}{|A|}$$

$$W(X; X_1, X_2, \ldots, X_N) = \frac{1}{2}\log\frac{\prod_{i=1}^{N} M_{ii}}{|A|^{N-1}}.$$

The proof is also obtained directly, through calculation.

A first generalization of the connection between the two types of measures will be made by considering function

$$\Delta^2(X_1, X_2, ..., X_N) = 1 - e^{-2W(X; X_1 X_2, ..., X_N)} \tag{23}$$

as an immediate passage, from the bidimensional case, presented by relation (18), to the N-dimensional case.

One can easily state that magnitude $\Delta^2 (X_1, X_2, ..., X_N)$ verifies the first six properties from theorem 5, obviously adapted to generalization.

The verification of the seventh property cannot be concerned, because, as we know, there is not a total indicator of dependence for the N-dimensional case.

Magnitude $\Gamma(X_1, X_2, ..., X_N)$ given by relation

$$\Gamma(X_1, X_2, ..., X_N) = \sqrt{1 - \frac{|A|^{N-1}}{\prod_{i=1}^{N} M_{ii}}} \tag{24}$$

is called total correlation coefficient of variables $X_1, X_2, ..., X_N$ with common distribution (21) (see also [5]). We notice immediately that, if $N = 2$, $\Gamma^2 = \rho^2$. We infer also the following theorem.

Theorem 8. If random vector $X = (X_1, X_2, ..., X_N)$ follows a distribution of type (18), then

$$\Delta^2(X_1, X_2, ..., XN) = \Gamma^2(X_1, X_2, ..., X_N). \tag{25}$$

Consequently, the seventh property from theorem 5 is also satisfied, a fact which leads to a generalization of theorem 5 and, at the same time, to a first generalization of Rényi's postulates (see also [5]). Magnitude $\Delta^2 (X_1, X_2, ..., X_N)$ defined by relation (23) is called total correlation coefficient of random variables $X_1, X_2, ..., X_N$) with a common distribution whatsoever. Consequently, we dispose also of a total correlation coefficient for the components of a random N-dimensional vector, differing generally from coefficient $\Gamma(X_1, X_2 ... X_N)$ which generalizes the classical correlation coefficient.

As a consequence of relation (25), for a certain distribution family, both total correlation coefficients are reciprocally and uniquely determined.

So as to present the second generalization of the connection between both measures, let us denote by:

$$\mathfrak{R}(X) = (\rho_{ij}),\ \rho_{ij} = \rho(x_i\ x_j),\ 1 \leqslant i,\ j \leqslant N \tag{26}$$

the correlation matrix, attached to random vector X, with

$$\mathfrak{D}(X) = (\delta_{ij}),\ \delta_{ij} = \delta(x_i\ x_j),\ 1 \leqslant i,\ j \leqslant N \tag{27}$$

the informational correlation coefficients given by (18), attached to random vector X, with

$$D(X) = (d_{ij}),\ d_{ij} = \begin{cases} \delta_{ij} & i = j \\ 0 & i \neq j \end{cases} \tag{28}$$

a diagonal matrix, attached to X, and with

$$U(X) = (u_{ij}),\ u_{ij} = 1\ \forall i, j = 1, 2, ..., N. \tag{29}$$

The following theorem is immediate.

Theorem 9. Matrix $\mathfrak{D}(X)$ defined by relation (27), satisfies the following properties:

1) it is defined for any random vector X, none of the components being constant with probability 1;

2) it is symmetrical, i.e. $\delta_{ij} = \delta_{ji}$ for any $1 \leqslant i, j \leqslant N$

3) $D(X) \leqslant \mathfrak{D}(X) \leqslant U(X)$ for any random vector X.

4) $\mathfrak{D}(X) = D(X)$ if any two components of random vector X are independent;

5) $\mathfrak{D}(X) = U(X)$ if between any two components of random vector X there is a functional dependence (this property is true only for the case in which there is no longer finiteness constraint for entropies);

6) if $\varphi(x) = (\varphi_1(x), \varphi_2(x), ..., \varphi_N(x))$ represents N real functions biunivocally defined on the real axis, then

$$\mathfrak{D}(\varphi(X)) = \mathfrak{D}(X)$$

7) if random vector X has distribution (21), then

$$\mathfrak{D}(X) = \mathfrak{R}(X).$$

It results, consequently, that we can have a matric indicator of dependence between the components of a random vector more powerful than the classical correlation matrix, an indicator which coincides with the matrix in the particular case of distribution families (21), to which also the normal distribution belongs.

We shall call matrix $\mathfrak{D}(X)$ given by relation (27), as informational correlation matrix attached to random vector X. Theorem 9 represents another generalization of theorem 5 and consequently, also for Rényi's postulates (see also [5]).

The two generalizations of the connection between the measure of interdependence and the classical correlation can be extended to their turn, for the more general indicators of intrdependence, presented by relation (10) and its particularizing for two vectors. We do not insist upon this problem.

REFERENCES

1. GUIASU S.: *Aplicații ale teoriei informației* (Applications of Information Theory). Ed. Academiei R.S.R. Bucharest, 1968.
2. LINFOOT E. H.: *An Informational Measure of Dependence*. Inf. and Control I, 1957, pp. 85–89.
3. PURCARU I.: *Asupra coeficientului de corelație informațională* (On the Informational Correlation Coefficient). Studii și cercetări matematide 3, 1976, pag. 343–350.

4. PURCARU I.: *Asupra unor măsuri ale interdependenței dintre două sau mai multe grupuri de variabile aleatoare* (On Some Measures of Interdependence between Two or More Groups of Random Variables). Studii și cercetări matematice 5, 1976, pag. 571—592.
5. PURCARU I.: *On Connection between the Statistical Correlation and the Informational Correlation.* Analele Universității București, 1976, pag. 15—19.
6. PURCARU I., POROJAN D.: *O metodă statistică-informațională de testare a interdependențelor unui sistem de factori* (A Statistical-Informational Method of Testing the Interdependence of a Factors System). Revista de statistică, 2, 1977, pag. 1—9
7. RENYI A.: *New Version of the Probabilistic Generalization of the Large Sieve.* Acta Mathematic Acad. Scie. Hung. vol. 10, 1959. pp. 217—226.
8. RENYI A.: *On Measure of Dependence.* Acta Math. Acad. Scie. Hung. Vol. 10, 1959, pp. 441—451.

SUR CERTAINES MESURES DE LA DEPENDANCE ENTRE LES VARIABLES ALEATOIRES

L'auteur présente la liaison qui existe entre deux mesures de la dépendance entre les variables aléatoires, suivie par deux généralisations de cette dépendance et dans le même temps de certaines extensions des postulats de Rényi ([7], [8]).

К НЕКОТОРЫМ МЕРАМ ЗАВИСИМОСТИ, СУЩЕСТВУЮЩЕЙ МЕЖДУ АЛЕАТОРНЫМИ ПЕРЕМЕННЫМИ

В настоящей работе указано на связь, существующая между двумя мерами зависимости алеаторных переменных, и приведены два обобщения этой связи, а также некоторые экстенсии постулатов Rényi ([7], [8]).

20.01.1978